

AMENDMENTS TO THE CLAIMS:

The listing of claims will replace all prior versions, and listings of claims in the application:

LISTING OF CLAIMS:

1. (Currently amended) A computer-implemented method of determining predictive models for a linked event detection system comprising the steps of:
 - determining source-identified training stories;
 - determining inter-story similarity vectors in a memory for at least one story-pair of the source-identified training stories;
 - determining link label information for the at least one story-pair, the link label information indicating the existence of at least one link between a pair of stories in the source-identified training stories and that the linked source-identified stories are related to the same event; and
 - determining and storing at least one predictive model in the memory based on the inter-story similarity vectors and the link label information.
2. (Original) The method of claim 1, wherein the step of determining inter-story similarity vectors comprises the steps of:
 - determining at least one inter-story similarity metric for the story-pairs; and
 - determining at least one source-pair statistics for the at least one story-pair.
3. (Original) The method of claim 2, wherein determining inter-story similarity vectors further comprise the step of normalizing the inter-story similarity metric based on the source-pair statistics.
4. (Original) The method of claim 2, wherein determining inter-story similarity vectors further comprise the step of incrementally normalizing the inter-story similarity metric based on the source-pair statistics.
5. (Original) The method of claim 2, wherein the inter-story similarity metric is

normalized based on at least one of subtraction and division.

6. (Original) The method of claim 2, wherein the inter-story similarity metric is at least one of a probability based similarity metric and a Euclidean based similarity metric.

7. (Original) The method of claim 6, wherein the probability based inter-story similarity metric is at least one of a Hellinger, a Tanimoto and a clarity distance based metric. (Original)

8. (Original) The method of claim 6, wherein the Euclidean based inter-story similarity metric is a cosine-distance based metric.

9. (Original) The method of claim 1, further comprising the step of transforming the source-identified training stories.

10. (Original) The method of claim 9, wherein transforming the source-identified training stories is at least one of translating, transcribing and linguistically transforming.

11. (Previously presented) The method of claim 2, wherein the inter-story similarity metrics are based on terms in at least one source-identified term frequency-inverse story frequency models.

12. (Original) The method of claim 11, wherein the terms in source-identified term frequency-inverse story frequency models are based on language.

13. (Original) The method of claim 11, wherein determining terms comprises the steps:

determining a reference language; and

determining reference language and non-reference language terms.

14. (Original) The method of claim 2, wherein the at least one inter-story similarity metric is normalized based on at least one of a source-pair identified

similarity statistic.

15. (Original) The method of claim 1, wherein the at least one predictive model is at least one of: a classifier, a support vector machine, a decision tree and a Naive-Bayes classifier.

16. (Original) The method of claim 2, wherein at least one of the source-pair similarity statistics are determined based on a source hierarchy.

17. (Original) The method of claim 16 wherein the source hierarchy is determined based on at least one source characteristic.

18. (Original) The method of claim 16 wherein the source characteristic is at least one of a language characteristic, an input mode characteristic, a genre characteristic, a source name characteristic and a transformation characteristic.

19. (Original) The method of claim 16 wherein the source-pair similarity statistic for a new source is determined based on at least one source characteristic of the new source.

20. (Currently amended) A linked event detection training system comprising:
an input/output circuit;
a memory;

a processor that receives source-identified training stories and associated link label information for at least one story-pair via the input/output circuit, the link label information indicating the existence of at least one link between a pair of stories in the source-identified training stories and that the linked source-identified stories are related to the same event;

an inter-story similarity vector determining circuit that determines inter-story similarity vectors in the memory for at least one story-pair of the source-identified training stories; and

a predictive model determining circuit that determines and stores at least one predictive model in the memory based on the inter-story similarity vectors

and the link label information.

21. (Original) The system of claim 20, wherein the inter-story similarity vector determining circuit is comprised of:

a similarity metric determining circuit that determines at least one inter-story similarity metric for the at least one story-pair; and

a similarity statistics determining circuit that determines at least one source-pair statistic for the at least one story-pair.

22. (Original) The system of claim 21, wherein the inter-story similarity vector determining circuit normalizes the inter-story similarity metric based on the source-pair statistics.

23. (Original) The system of claim 21, wherein the inter-story similarity vector determining circuit incrementally normalizes the inter-story similarity metric based on the source-pair statistics.

24. (Original) The system of claim 21, wherein at least one of the inter-story similarity metrics is normalized based on at least one of a subtraction and a division operation.

25. (Original) The system of claim 21, wherein at least one of the inter-story similarity metrics is at least one of a probability based similarity metric and a Euclidean based similarity metric.

26. (Original) The system of claim 25, wherein the probability based inter-story similarity metric is at least one of a Hellinger, a Tanimoto and a clarity distance based metric.

27. (Original) The system of claim 25, wherein the Euclidean based inter-story similarity metric is a cosine-distance based metric.

28. (Original) The system of claim 20, wherein the source-identified training stories are transformed.

29. (Original) The system of claim 28, wherein transforming the source-identified training stories is at least one of translating, transcribing and linguistically transforming.

30. (Original) The system of claim 20, wherein the inter-story similarity metrics are based on terms in at least one source-identified term frequency-inverse story frequency model.

31. (Original) The system of claim 30, wherein the terms in the source-identified term frequency-inverse story frequency models are based on language.

32. (Original) The system of claim 30, wherein the processor determines terms based on a reference language; and determining reference language and non-reference language terms.

33. (Original) The system of claim 21 wherein the at least one inter-story similarity metric is normalized based on at least one of a source-pair identified similarity statistic.

34. (Original) The system of claim 20, wherein the at least one predictive model is at least one of: a classifier, a support vector machine, a decision tree and a Naive-Bayes classifier.

35. (Original) The system of claim 21, wherein the source-pair identified similarity statistic is determined based on a source hierarchy.

36. (Original) The system of claim 35, wherein the source hierarchy is determined based on at least one of a source characteristic.

37. (Original) The system of claim 35, wherein the source characteristic is at least one of a language characteristic, an input mode characteristic, a genre characteristic, a source name characteristic and a transformation characteristic.

38. (Original) The system of claim 35, wherein the source-pair similarity statistic for a new source is determined based on at least one source characteristics

of the new source.

39. (Currently amended) A computer-implemented method of linked event detection comprising the steps of:

determining source-identified stories;

determining inter-story similarity vectors in a memory for the story-pairs of the source-identified stories;

determining at least one predictive model in ~~a~~the memory for link detection;
and

determining a link between the story-pairs based on the predictive model and the inter-story similarity vector; and

~~indicating~~displaying the link on a computer or storing the link in an
information repository, the link indicating the story-pairs are related to the
same event.

40. (Previously presented) The method of claim 39, wherein the step of determining inter-story similarity vectors comprises the steps of:

determining at least one inter-story similarity metric for each story-pair; and

determining source-pair statistics for the story-pairs.

41. (Original) The method of claim 40, wherein determining inter-story similarity vectors further comprise the step of normalizing the inter-story similarity metric based on the source-pair statistics.

42. (Original) The method of claim 40, wherein determining inter-story similarity vectors further comprise the step of incrementally normalizing the inter-story similarity metric based on the source-pair statistics.

43. (Original) The method of claim 40, wherein the inter-story similarity metric is normalized based on at least one of subtraction and division.

44. (Original) The method of claim 40, wherein the inter-story similarity metric

is at least one of a probability based similarity metric and a Euclidean based similarity metric.

45. (Original) The method of claim 44, wherein the probability based inter-story similarity metric is at least one of a Hellinger, a Tanimoto and a clarity distance based metric.

46. (Original) The method of claim 44, wherein the Euclidean based similarity metric is a cosine-distance based metric.

47. (Original) The method of claim 39, further comprising the step of transforming the source-identified training stories.

48. (Original) The method of claim 47, wherein transforming the source-identified training stories is at least one of translating, transcribing and linguistically transforming.

49. (Previously presented) The method of claim 40, wherein the inter-story similarity metrics are based on terms in at least one source-identified term frequency-inverse story frequency models.

50. (Original) The method of claim 49, wherein the terms in source-identified term frequency-inverse story frequency models are based on language.

51. (Original) The method of claim 49, wherein determining terms comprises the steps:

determining a reference language; and

determining reference language and non-reference language terms.

52. (Original) The method of claim 40, wherein the at least one inter-story similarity metric is normalized based on at least one of a source-pair identified similarity statistic.

53. (Original) The method of claim 39, wherein the at least one predictive

model is at least one of: a classifier, a support vector machine and a decision tree, a Naive-Bayes-classifier.

54. (Original) The method of claim 40, wherein the source-pair identified similarity statistic is determined based on a source hierarchy.

55. (Original) The method of claim 54, wherein the source hierarchy is determined based on at least one of a source characteristic.

56. (Original) The method of claim 54, wherein the source characteristic is at least one of a language characteristic, an input mode characteristic, a genre characteristic, a source name characteristic and a transformation characteristic.

57. (Original) The method of claim 54, wherein the source-pair similarity statistic for a new source is determined based on at least one source characteristics of the new source.

58. (Currently amended) A linked event detection system comprising:

an input/output circuit;

a memory;

a processor that receives source-identified stories via the input/output circuit;

an inter-story similarity vector determining circuit that determines inter-story similarity vectors in the memory for the story-pairs of the source-identified stories; and

a link determining circuit that determines and indicates-displays on a computer or stores in an information repository, links between story-pairs based on a predictive model in the memory and the inter-story similarity vectors, the links indicating the story-pairs are related to the same event.

59. (Currently amended) The method-system of claim 58, wherein the inter-story similarity vector determining circuit is comprised of:

a similarity metric determining circuit that determines at least one inter-story similarity metric for the story-pairs; and

a similarity statistics determining circuit that determines source-pair statistics for the story-pairs.

60. (Original) The system of claim 59, wherein the inter-story similarity vector determining circuit normalizes the inter-story similarity metric based on the source-pair statistics.

61. (Original) The system of claim 59, wherein the inter-story similarity vector determining circuit incrementally normalizes the inter-story similarity metric based on the source-pair statistics.

62. (Original) The system of claim 59, wherein at least one of the inter-story similarity metrics is normalized based on at least one of a subtraction and a division operation.

63. (Original) The system of claim 59, wherein at least one of the inter-story similarity metrics is at least one of a probability based similarity metric and a Euclidean based similarity metric.

64. (Original) The system of claim 63, wherein the probability based inter-story similarity metric is at least one of a Hellinger, a Tanimoto and a clarity distance based metric.

65. (Original) The system of claim 63, wherein the Euclidean based inter-story similarity metric is a cosine-distance based metric.

66. (Original) The system of claim 58, wherein the source-identified training stories are transformed.

67. (Original) The system of claim 66, wherein transforming the source-identified training stories is at least one of translating, transcribing and linguistically transforming.

68. (Previously presented) The system of claim 59, wherein the inter-story similarity metrics are based on terms in at least one source-identified term

frequency-inverse story frequency model.

69. (Original) The system of claim 68, wherein the terms in the source-identified term frequency-inverse story frequency models are based on language.

70. (Original) The system of claim 68, wherein the processor determines terms based on a reference language; and non-reference language terms.

71. (Original) The system of claim 59, wherein the at least one inter-story similarity metric is normalized based on at least one of a source-pair identified similarity statistic.

72. (Original) The system of claim 58, wherein the predictive model is at least one of: a classifier, a support vector machine and a decision tree, a Naive-Bayes classifier.

73. (Original) The system of claim 59, wherein the source-pair identified similarity statistic is determined based on a source hierarchy.

74. (Original) The system of claim 73, wherein the source hierarchy is determined based on at least one of a source characteristic.

75. (Original) The system of claim 73, wherein the source characteristic is at least one of a language characteristic, an input mode characteristic, a genre characteristic, a source name characteristic and a transformation characteristic.

76. (Original) The system of claim 73, wherein the source-pair similarity statistic for a new source is determined based on at least one source characteristics of the new source.

77. (Currently amended) A method of determining a stopword list comprising the steps of:

determining a source-identified training corpus of text information;

determining a verified first source-mode transformation of the source-

identified training corpus text from a first ~~source~~-mode to a second ~~source~~ mode based on at least one of a verified transcription and a verified translation;

determining an un-verified second source-mode transformation of the source-identified training corpus text from a first ~~source~~-mode to a second ~~source~~ mode;

determining at least one transformation ~~errors~~-error associated with distribution differences between the first and second transformations and identified sources;

determining and storing at least one source-specific transformation ~~actions~~ action for the determined transformation errors in a memory; and

identifying and transforming transformation errors in other transformed source-identified texts based on the source-specific transformation actions in the memory.

78. (Currently amended) The method of claim 77, wherein the first ~~source~~ mode is at least one of a text source, an optical character recognition source and an automatic speech recognition source.

79. (Currently amended) The method of claim 77, wherein the second ~~source~~ mode is at least one of a text source, an optical character recognition source and an automatic speech recognition source.

80. (Original) The method of claim 77, wherein the source-specific transformation is at least one of a removal, a repair and a normalization transformation.

81. (Currently amended) Computer readable storage medium comprising: computer readable program code embodied on the computer readable storage medium, the computer readable program code ~~executable~~-processable to program a computer to determine at least one predictive model for a linked event detection system by executing steps ~~comprising the steps of~~:

determining source-identified training stories;

determining inter-story similarity vectors in a memory for at least one story-pair;

determining link label information for the at least one story-pair of the source-identified training stories, the link label information indicating training stories related to the same event; and

determining and storing at least one predictive model in the memory based on the inter-story similarity vectors and the link label information.

82. (Currently amended) Computer readable storage medium comprising: computer readable program code embodied on the computer readable storage medium, the computer readable program code ~~usable~~ processable to program a computer to determine at least one predictive model for a linked event detection system, the computer readable program code comprising:

instructions to determine source-identified training stories;

instructions to determine inter-story similarity vectors in a memory for at least one story-pair of the source-identified training stories;

instructions to determine link label information for the at least one story-pair, the link label information indicating training stories related to the same event; and

instructions to determine and store at least one predictive model in the memory based on the inter-story similarity vectors and the link label information.

83. (Currently amended) Computer readable storage medium comprising: computer readable program code embodied on the computer readable storage medium, the computer readable program code ~~executable~~ processable to program a computer to detect linked events by executing steps comprising ~~the steps of~~:

determining source-identified stories;

determining inter-story similarity vectors in a memory for the at least one story-pair of the source-identified stories;

determining at least one predictive model in the memory for link detection;

and

determining a link between story-pairs based on the at least one predictive model and the inter-story similarity vectors, the link indicating the story-pairs are related to the same event; and

indicating-displaying the link on a computer or storing the link in an information repository.

84. (Currently amended) Computer readable storage medium comprising: computer readable program code embodied on the computer readable storage medium, the computer readable program code executable-processable to program a computer to detect linked events, the computer readable program code comprising the steps of:

instructions to determine source-identified stories;

instructions to determine inter-story similarity vectors in a memory for the at least one story-pair of the source-identified stories;

instructions to determine at least one predictive model in a-the memory for link detection;

instructions to determine a link between story-pairs based on the predictive model and the inter-story similarity vectors, the link indicating the story-pairs are related to the same event; and

instructions to indicate-display the link on a computer or store the link in an information repository.

85. (Original) The method of claim 2, wherein determining at least one source-pair statistic for the at least one story-pair is based on at least one of a similarity metric and a statistic associated with the metric.

86. (Original) The system of claim 21, wherein determining at least one source-pair statistic for the at least one story-pair is based on at least one of a similarity metric and a statistic associated with the metric.

87. (Original) The method of claim 39, wherein at least one of the predictive

models is a trained predictive model.

88. (Original) The system of claim 58, wherein at least one of the predictive models is a trained predictive model.